

DEFINIÇÃO DE LIMITES DE VELOCIDADE EM ESTRADAS INTERURBANAS – ABORDAGEM ENVOLVENDO ÁRVORES DE DECISÃO

Nuno Gregório¹, Ana Bastos Silva² e Alvaro Seco²

¹ Universidade de Coimbra, Departamento de Engenharia Civil, Centro de Investigação do Território, Transportes e Ambiente (CITTA), Rua Luís Reis Santos – Pólo II, 3030-788 Coimbra, Portugal

email: nuno.gregorio@uc.pt <https://citta.fe.up.pt/>

² Universidade de Coimbra, Departamento de Engenharia Civil, Centro de Investigação do Território, Transportes e Ambiente (CITTA), Rua Luís Reis Santos – Pólo II, 3030-788 Coimbra, Portugal

Sumário

O presente trabalho pretendeu desenvolver uma metodologia de apoio à decisão, que de forma acessível e generalizada, potencie a definição do limite de velocidade em estradas nacionais, atravessando vários tipos de ambiente rodoviário. Para tal, foi considerado um conjunto de fatores relacionados com diversos aspetos da via e, nomeadamente, com a sua integração na envolvente. O sistema desenvolvido é um sistema especialista, replicando o processo decisório de um grupo de especialistas, tendo sido utilizadas árvores de decisão. A calibração envolveu a recolha de dados numa extensão superior a 300 km de estradas nacionais, de diversas regiões.

Palavras-chave: Velocidade Máxima; Ambiente Rodoviário; Sistema Especialista; Árvores de Decisão

1 INTRODUÇÃO

Durante a segunda metade do século passado, o intenso desenvolvimento urbano em várias regiões do Mundo levou ao surgimento de áreas urbanizadas com um padrão de cariz mais disperso e não-consolidado, em contraste com o povoamento mais concentrado, que fora até então a regra nos aglomerados populacionais. Este tipo de situações, que ocorrem frequentemente ao longo de eixos viários pré-existentes, é relativamente comum em regiões em desenvolvimento, ou que há pouco atingiram o estado do pleno desenvolvimento, como é o caso de várias regiões portuguesas – nomeadamente, no Norte e no Centro do País. Uma das principais consequências que advém deste fenómeno é a maior dificuldade (ou, mesmo, a impossibilidade) de definição, de forma clara, do limite entre as zonas urbanas e os territórios puramente rurais. Frequentemente, esta fronteira corresponde a uma zona mais alargada, a zona de transição, que apresenta características mistas, conferindo-lhe uma natureza suburbana ou periurbana.

Esta situação tem levado ao surgimento, cada vez mais comum, de ambientes rodoviários complexos ao longo dos principais eixos rodoviários e das suas áreas envolventes. Nestas zonas, os princípios tradicionalmente seguidos na conceção e na gestão rodoviárias têm-se tornado insuficientes e ineficazes. Um caso particular desta problemática é a definição do limite de velocidade. Em primeiro lugar, refira-se a inexistência de uma metodologia consensualmente reconhecida e adotada pelas comunidades científica e técnica nos vários contextos, bem como em sede regulamentar [1]. Por outro lado, em países como Portugal, uma parte substancial do processo decisório é confiado à gestão local, gerando evidentes problemas de coerência e de homogeneidade no conjunto da rede rodoviária [2-4]. Frequentemente, a rede é encarada de um ponto de vista dual, subdivida apenas em vias urbanas e estradas rurais, deliberadamente ignorando a existência das zonas de transição, às quais é frequentemente associado o regime urbano de circulação. Além disso, no caso específico português, o habitual uso de sinais de identificação de localidade – eles próprios normalmente implementados de forma arbitrária e pouco consistente – constitui um fator agravante deste problema.

Ao longo das últimas décadas, várias metodologias têm sido concebidas e adotadas para a definição do limite de velocidade em vias interurbanas. Os métodos mais comumente utilizados integram-se na família dos Métodos

da Engenharia, que se baseia nas características geométricas do traçado, destacando-se o Método da Velocidade Operacional. Este método utiliza o conceito de percentil 85 da velocidade (V85), estabelecendo o limite de velocidade num valor igual ou superior a este parâmetro, que é posteriormente ajustado de acordo com as condições de tráfego e as características infraestruturais da via [1]. Este procedimento procura atender ao facto de o risco mínimo de acidente para o condutor ocorrer quando a velocidade é, aproximadamente, coincidente com um valor correspondente a um desvio-padrão somado à média da distribuição de velocidades, num valor que é próximo do V85 [5]. Consequentemente, esta metodologia contribui para um regime de velocidades mais uniforme, na medida em que o limite de velocidade corresponde à percepção da maioria dos condutores acerca da velocidade de circulação mais adequada [1,6]. Contudo, este método tem sido criticado com base no seu irrealismo, na sua falta de racionalidade (causada pela percepção muitas vezes errónea do condutor acerca dos impactos da velocidade), na sua falta de eficácia em ambientes residenciais ou na forma inadequada com que trata a consistência da velocidade; sendo ainda dado como responsável por um aumento gradual da velocidade média de circulação [7-9].

Outras abordagens incluem as metodologias da Minimização do Dano e da Otimização Económica. A primeira destas estratégias baseia-se em imperativos éticos, considerando a tolerância e integridade humanas em situações de sinistro, tendo sido criticada devido ao seu irrealismo e à falta de racionalidade e de credibilidade, do ponto de vista do condutor [10-12]. Por outro lado, os métodos de Otimização Económica focam-se na atribuição de um valor monetário a todos os custos relacionados com a mobilidade, sendo esta informação normalmente difícil de obter [13].

Por sua vez, uma nova família de metodologias, utilizando sistemas especialistas, tem sido desenvolvida e implementada nalguns países, sendo os casos mais significativos os da Austrália e dos Estados Unidos. Estes métodos procuraram solucionar a inconsistência e falta de uniformidade características dos métodos da engenharia, simulando e reproduzindo o comportamento de especialistas no processo decisório de definição da velocidade máxima [14,15]. Em geral, estes sistemas têm produzido resultados mais consistentes, fiáveis e abrangentes, sendo ainda mais facilmente reproduzíveis em contextos distintos [1].

Desta forma, uma nova metodologia foi desenvolvida para o apoio à decisão na definição do limite de velocidade em estradas interurbanas de faixa de rodagem única (designadamente, vias comumente classificadas em Portugal como estradas nacionais e regionais), atravessando ambientes rodoviários diversos. Esta metodologia baseou-se em fatores relacionados com o ambiente rodoviário prevalecente, bem como com a área envolvente à via, procurando utilizar informação facilmente recolhida, através de fontes remotas, de modo a garantir a sua acessibilidade. Este modelo é um sistema especialista, tendo implicado a constituição de uma base de dados contendo informação recolhida ao longo de mais de 300 km de estradas nacionais, nos dois sentidos, em três diferentes regiões de Portugal (Norte, Centro e Sul), atravessando áreas urbanas, rurais e de transição. Esta base de dados envolveu ainda a recolha da avaliação de um grupo de especialistas em Engenharia de Tráfego, relativamente ao limite de velocidade a implementar em cada trecho analisado.

Este trabalho integra-se numa linha de investigação, na qual se incluem [16] e [17], tendo recorrido a uma base de dados mais abrangente – em termos de extensão de via e, como tal, de número de secções incluídas; em termos de diversidade de regiões e de ambientes rodoviários atravessados; e no que respeita ao conjunto de variáveis explicativas, estando estas relacionadas com as características funcionais da via, o ambiente rodoviário e a envolvente. Além disso, foi implementado um novo método de realização da avaliação por parte dos especialistas, tal como descrito em [18] e [19], decorrendo esta em laboratório, em condições mais uniformes e estáveis, permitindo reforçar a sua replicabilidade. Embora a utilização de fatores relacionados com as características geométricas da via na atribuição da velocidade máxima tenha sido extensivamente estudada, procurou-se colocar o presente trabalho noutra enquadramento, preenchendo uma lacuna no estado atual do conhecimento. Não obstante, a influência das características geométricas da via deverá ser considerada em fases posteriores, analisando secções mais longas, e enquadrada do ponto de vista da homogeneidade de traçado e da consistência das velocidades.

Anteriormente, a atribuição do limite de velocidade foi encarada como sendo um problema de escolha discreta, tendo sido utilizados modelos do tipo *logit*. No presente trabalho, foi adotada uma nova metodologia envolvendo árvores de decisão. Para tal, foram testados e comparados dois algoritmos, dos mais simples desta família: os métodos CART (*Classification and Regression Trees*) e C4.5. Numa fase posterior desta linha de investigação, serão aplicados métodos mais complexos. O modelo final é uma árvore podada do tipo C4.5, tendo a sua estrutura interna e os pressupostos nos quais se baseia sido analisados e discutidos.

2 SISTEMAS ESPECIALISTAS

Os sistemas especialistas (*expert systems*) são programas computacionais que procuram reproduzir o conhecimento e os processos de decisão humanos de forma artificial, através de um conjunto complexo de regras e relações. Mais especificamente, um sistema deste tipo recorre a uma base de dados para, através de algoritmos de decisão, resolver problemas complexos, permitindo simular e reproduzir o comportamento, o raciocínio, a avaliação e os processos decisórios de especialistas da área do conhecimento em questão [15]. Em geral, estes sistemas têm sido reconhecidos pela sua abrangência, consistência, sistematicidade e reprodutibilidade em contextos diversos [1,14].

A componente fundamental deste tipo de sistemas é a base de dados que lhe serve de suporte, designada, neste contexto, como base de conhecimento (*knowledge base*). É nesta base que o sistema assenta o seu processo decisório, pelo que a sua constituição e estruturação é crucial para assegurar a precisão e consistência dos resultados. Para tal, a base de conhecimento deve conter informação proveniente de um conjunto de especialistas, representativos da comunidade técnica e científica, relativa à sua experiência e conhecimento específicos, sendo organizada e estruturada em tarefas a executar e decisões a tomar [7]. Por outro lado, o algoritmo de decisão, a função do sistema, pode consistir numa vasta gama de funções, desde preditores simples até algoritmos mais complexos de inteligência artificial, como redes neuronais ou modelos de lógica difusa (*fuzzy*) [20].

Os sistemas especialistas têm um longo historial de aplicação em problemas da Engenharia de Transportes, nomeadamente no planeamento de redes de transportes, na segurança rodoviária, no controlo e na gestão do tráfego rodoviário ou no controlo do tráfego aéreo [21-23]. Concretamente neste campo, os casos mais significativos de aplicação têm sido a Austrália, com a implementação dos sistemas da família XLIMITS, e os Estados Unidos da América, com o USLIMITS (atualmente, já numa segunda fase de desenvolvimento, o USLIMITS2). No primeiro caso, o processo de desenvolvimento iniciou-se ainda nos anos 80 do século passado no estado de Victoria, tendo mais tarde sido alargado aos restantes estados que formam a Austrália [14]. Estes sistemas utilizam um programa do tipo *hard-coded*, em cuja aplicação o utilizador insere informação – quer informação objetiva, quer subjetiva – acerca das características da secção analisada, sendo o resultado final determinado através de uma árvore de decisão [14,24]. Por outro lado, o sistema USLIMITS2 foi desenvolvido a partir da experiência australiana, tendo envolvido um conjunto de especialistas repartido em dois grupos, cada um deles intervindo numa fase distinta: um primeiro grupo participou na conceção do sistema; o outro grupo interveio ao nível da recolha de informação, permitindo posteriormente estabelecer o conjunto de regras de aplicação que se encontram subjacentes à implementação do sistema [24,25]. Neste caso, o sistema solicita, *a priori*, a introdução do tipo de via (de estrada rural até via rápida urbana), sendo o resultado final quanto ao valor do limite de velocidade acompanhado de recomendações técnicas e alertas de implementação [1].

3 ABORDAGEM METODOLÓGICA

3.1 Base de conhecimento

A base de conhecimento é a componente central do sistema especialista, permitindo estabelecer a relação entre os fatores explicativos incluídos no modelo e a informação introduzida pelos especialistas, bem como a sua posterior replicação noutras vias.

Os dados relativos a cada um destes fatores, bem como a avaliação dos especialistas em relação ao limite de velocidade, foram recolhidos ao longo de 7 itinerários, abrangendo trechos de estradas nacionais de faixa de rodagem única, totalizando 304,2 km. O processo foi, sucessivamente, levado a cabo em ambos os sentidos. Estes itinerários foram selecionados tendo a variabilidade das suas secções como principal critério, designadamente em dois aspetos:

- Variabilidade *intra-secção*: itinerários atravessando ambientes rodoviários distintos, nomeadamente áreas urbanas, zonas de transição e áreas rurais;
- Variabilidade *inter-secção*: itinerários em regiões com diferentes tipos de ocupação urbana, nomeadamente ocupação dispersa e difusa (casos do Norte e do Centro) e com uma tipologia urbana concentrada e consolidada (caso do Sul).

Cada um destes itinerários foi, por sua vez, subdividido em secções com uma extensão variável (mas nunca superior a 700 metros). A definição destas secções foi efetuada de forma a que as mesmas correspondessem a segmentos com características o mais uniformes e homogéneas possível, permitindo, desta forma, uma posterior avaliação mais estável e consistente por parte dos especialistas. Os itinerários considerados foram subdivididos num total de 1036 secções.

O grupo de especialistas interveniente neste processo, composto por quatro elementos, foi selecionado atendendo à sua experiência e conhecimento técnicos e científicos. Cada um destes especialistas procedeu à avaliação de cada uma das secções em análise de forma independente face às adjacentes, atribuindo a cada uma delas, sucessivamente, um valor para o respetivo limite de velocidade. Para tal, foram contempladas três alternativas:

- 50 km/h, o limite de velocidade regulamentar para vias em zona urbana, em Portugal;
- 70 km/h, um valor característico das zonas de transição;
- 90 km/h, o limite de velocidade regulamentar para estradas rurais, em Portugal.

Neste processo de avaliação, foi solicitado a cada um dos especialistas que tivesse em conta apenas a influência dos elementos funcionais da via e da sua interação com o ambiente envolvente. Os limites de velocidade em vigor ao longo de cada uma das secções não foram tidos em consideração. O procedimento apresentado em [18] e [19] foi adotado, tendo os especialistas efetuado a sua avaliação em laboratório. Para tal, cada especialista assistiu a gravações em vídeo de cada um dos itinerários, em ambos os sentidos, emitindo o seu juízo relativo à velocidade máxima admissível para cada secção, de forma sucessiva. A adoção deste procedimento procurou assegurar condições o mais homogéneas possível para cada um dos ensaios, garantindo um nível adequado de replicabilidade, dado o ambiente mais controlado [18]. No domínio da Segurança Rodoviária, a utilização de abordagens conservadoras deve ser privilegiada face ao emprego de valores médios. Consequentemente, um especialista virtual conservador, tal como definido em [17], foi adotado, implicando a seleção do segundo valor mais restritivo do limite de velocidade em cada secção, de entre as várias propostas emitidas pelos especialistas.

As variáveis explicativas incluídas na base de conhecimento encontram-se representadas no Quadro 1. Estas variáveis são sobretudo representativas das características funcionais da via, do ambiente rodoviário e da envolvente. Os dados relativos a cada uma destas variáveis foram recolhidos através da observação e manipulação de imagem aérea e de satélite, bem como da exibição dos vídeos anteriormente referidos. No Quadro 1, as variáveis de carácter binário representam a existência da característica descrita, enquanto que as variáveis contínuas exprimem a proporção da extensão da secção em que essa característica ocorre. Outras variáveis são expressas em termos de densidade de ocorrências (número de ocorrências por 100 metros), algo que advém do facto de a extensão das secções ser variável. As variáveis relativas à densidade de edificações (EDR e EDL) e à distância do edificado à faixa de rodagem (MDR, MDL, MEDR e MEDL) foram determinadas contemplando apenas as edificações contidas numa banda até 30 metros da faixa de rodagem, seguindo o estabelecido em [17]. Além disso, nestas variáveis relacionadas com distância, apenas as edificações que formam o alinhamento frontal do edificado – isto é, das edificações que têm contacto direto com a faixa de rodagem – foram contabilizadas.

3.2 Árvores de decisão

Os algoritmos da família das árvores de decisão são preditores não-paramétricos de aprendizagem supervisionada, em que os dados são sucessivamente classificados numa estrutura de forma arbórea, correspondendo cada nó (*folha*) a uma função e a uma classe de instâncias. Estes nós, quando intermédios, apresentam uma conjunção condicional que divide o domínio de uma dada variável, interligando-se em *ramos*, em partições binárias recursivas do conjunto de fatores [26,27] – ou seja, cada nó dividir-se-á noutros dois nós, exceto se for nó terminal. A variável e a localização da partição são determinadas em função da minimização da impureza do nó (e, consequentemente, da maximização da homogeneidade), o que poderá ser efetuado através de critérios como o erro de classificação, o índice de Gini ou o índice de entropia [27].

Um crescimento excessivo da árvore poderá levar a sobreajustamento (*overfitting*) e a perda de qualidade do preditor numa validação efetuada com um conjunto de dados independente. Esta situação é, normalmente, abordada em duas fases: em primeiro lugar, a árvore é desenvolvida até atingir uma dimensão excessiva; posteriormente, é podada (*pruning*), até atingir uma dimensão ótima, determinada com base no conjunto independente de dados de validação [27].

Quadro 1. Variáveis explicativas incluídas na base de dados

Variável	Descrição	Unidade	
EDR	Densidade de edificações, lado adjacente	Elementos/100 metros	
EDL	Densidade de edificações, lado oposto		
MEDR	Distância média entre o edificado e a faixa de rodagem, lado adjacente	Metros	
MEDL	Distância média entre o edificado e a faixa de rodagem, lado oposto		
MDR	Distância mínima entre o edificado e a faixa de rodagem, lado adjacente		
MDL	Distância mínima entre o edificado e a faixa de rodagem, lado oposto		
ROUND	Rotunda	Binária	
INTCR	Cruzamento com 4 ou mais acessos		
INTENR	Entroncamento com 3 acessos, lado adjacente	Elementos/100 metros	
INTENL	Entroncamento com 3 acessos, lado oposto		
NATERR	Acesso local, lado adjacente		
NATERL	Acesso local, lado oposto		
AGROR	Via rural, lado adjacente		
AGROL	Via rural, lado oposto		
LOGR	Acesso a logradouro público, incluindo estacionamento, lado adjacente		
LOGL	Acesso a logradouro público, incluindo estacionamento, lado oposto		
ESTPR	Área aberta de estacionamento informal, lado adjacente		
ESTPL	Área aberta de estacionamento informal, lado oposto		
NGASR	Posto de abastecimento de combustível, lado adjacente	Binária	
NGASL	Posto de abastecimento de combustível, lado oposto		
NAPARR	Lugares formais de estacionamento, lado adjacente	Elementos/100 metros	
NAPARL	Lugares formais de estacionamento, lado oposto		
NIPARR	Lugares informais de estacionamento, lado adjacente		
NIPARL	Lugares informais de estacionamento, lado oposto		
GARAGR	Acesso a estacionamento privado, incluindo garagens, lado adjacente		
GARAGL	Acesso a estacionamento privado, incluindo garagens, lado oposto		
NAPEHR	Acesso pedonal privado, através de edificado, lado adjacente		
NAPEHL	Acesso pedonal privado, através de edificado, lado oposto		
NAPEDR	Acesso pedonal público, lado adjacente		
NAPEDL	Acesso pedonal público, lado oposto		
NCRO	Passagem pedonal	Elementos/100 metros	
NCROS	Passagem pedonal controlada através de semáforos		
NBUSR	Paragem de transportes públicos, lado adjacente		
NBUSL	Paragem de transportes públicos, lado oposto		
NBUSPR	Paragem de transportes públicos segregada, lado adjacente		
NBUSPL	Paragem de transportes públicos segregada, lado oposto		
SIGN	Sinais luminosos de controlo de velocidade		Binária
SIDEWR	Passeio, lado adjacente		Proporção
SIDEWL	Passeio, lado oposto		
ISL	Ilha central com marcação rodoviária		
ISLPH	Ilha central com separador físico		
LELAN	Via de viragem à esquerda		
RILAN	Via de viragem à direita		
DLANR	Vias múltiplas, no sentido do movimento		
DLANL	Vias múltiplas, no sentido oposto		
SHLDR	Berma, lado adjacente		
SHLDL	Berma, lado oposto		
RAILR	Guarda de segurança, lado adjacente		
RAILL	Guarda de segurança, lado oposto		
VEDR	Vedação na berma, lado adjacente		
VEDL	Vedação na berma, lado oposto		
OPACR	Objeto longo na berma, incluindo muros, lado adjacente		
OPACL	Objeto longo na berma, incluindo muros, lado oposto		

Um dos algoritmos envolvendo árvores de decisão mais usados é o modelo conhecido como CART (*Classification and Regression Trees*), definido por [28], que estabeleceu e utilizou, pela primeira vez, vários destes conceitos. Este modelo tem sido aplicado de forma generalizada a vários tipos de problemas, em vários domínios. Outro algoritmo, apresentado por [29], define as árvores de decisão do tipo C4.5. Este modelo recorre a um critério de minimização da impureza baseado na entropia, sendo a poda aplicada com base na redução do erro estimado.

Ambos os algoritmos foram aplicados à base de dados, nas suas versões não-podada e podada, tendo sido posteriormente comparados. A validação de cada modelo foi efetuada com base num conjunto independente de instâncias gerado aleatoriamente a partir da base de dados, correspondendo a 30% do seu total. Neste processo, foi utilizado o *software* WEKA 3.8 (© University of Waikato, Nova Zelândia).

4 RESULTADOS

As árvores obtidas pela aplicação de cada um dos algoritmos, nas suas versões antes e após a poda, foram analisadas e comparadas. Os dados relativos ao número de instâncias corretamente classificadas e à raiz quadrada do erro quadrático médio (RMSE) – uma medida da precisão do ajustamento aos dados – de cada uma das árvores são apresentados no Quadro 2.

Quadro 2. Instâncias corretamente classificadas e raiz do erro quadrático médio (RMSE) em cada árvore

	CART não-podada	CART podada	C4.5 não-podada	C4.5 podada
Corretamente Classificadas	67,2%	69,5%	72,7%	73,6%
RMSE	0,444	0,405	0,413	0,396

A análise destes indicadores permite extrair algumas conclusões preliminares. Em primeiro lugar, e tal como era expectável, ambas as árvores assumem um melhor desempenho na sua versão podada. De facto, esta operação permite obter uma dimensão ótima da árvore, evitando a redundância de parte da estrutura do preditor. No que respeita ao algoritmo, as árvores obtidas através da aplicação do C4.5 apresentam um melhor ajustamento aos dados que as estimadas através do modelo CART. Apesar de a diferença ao nível da RMSE não ser significativa, o algoritmo C4.5 alcançou um melhor desempenho, quer quanto a este indicador (valores mais reduzidos antes e após a poda), quer quanto à classificação das secções que formam parte do conjunto de validação, tendo obtido uma classificação correta em perto de 75%. Deste modo, fica patente que as árvores do tipo C4.5 são, à partida, mais precisas enquanto preditor.

A análise das matrizes de confusão permite avaliar, de uma forma mais exaustiva, o desempenho destes algoritmos. Nestas matrizes, são comparadas, para cada uma das categorias, as instâncias corretamente classificadas com as que foram incorretamente atribuídas a uma das outras categorias. As matrizes de confusão relativas a cada um dos quatro casos acima descritos são apresentadas no Quadro 3, para as árvores do tipo CART, e no Quadro 4, para as árvores C4.5.

Quadro 3. Matrizes de confusão de cada árvore CART

	CART não-podada			CART podada		
	67,2%			69,5%		
Classificada como →	50 km/h	70 km/h	90 km/h	50 km/h	70 km/h	90 km/h
50 km/h	77,1%	22,9%	0%	81,7%	18,3%	0%
70 km/h	27,59%	59,1%	13,4%	26,8%	58,3%	14,9%
90 km/h	3,8%	33,9%	62,3%	1,9%	32,1%	66%

Quadro 4. Matrizes de confusão de cada árvore C4.5

Corretamente Classificadas	C4.5 não-podada			C4.5 podada		
	72,7%			73,6%		
Classificada como →	50 km/h	70 km/h	90 km/h	50 km/h	70 km/h	90 km/h
50 km/h	85,5%	14,5%	0%	87,8%	12,2%	0%
70 km/h	18,9%	68,5%	12,6%	18,1%	68,5%	13,4%
90 km/h	3,8%	45,3%	50,9%	1,9%	47,2%	50,9%

Como ponto prévio, há que referir que, em qualquer dos casos, as ocorrências menos frequentes são aquelas que correspondem à situação mais gravosa, ou seja, à atribuição de 90 km/h a secções de 50 km/h e vice-versa. Este facto é um dado positivo da aplicação de ambos os algoritmos pois, quer do ponto de vista do desempenho dos preditores, quer na perspetiva da Segurança Rodoviária, essas situações configuram uma maior discrepância entre o resultado do preditor e a realidade. Por outro lado, ambos os algoritmos se comportam significativamente melhor na alternativa dos 50 km/h, em que são alcançados níveis elevados de precisão, o que se poderá dever ao facto de estas secções constituírem o caso mais numeroso resultante da avaliação dos especialistas. Já na alternativa com uma pior precisão, observa-se uma divergência entre os algoritmos. Enquanto que a alternativa dos 70 km/h apresenta um pior desempenho nas árvores do tipo CART, já as árvores C4.5 exibem uma precisão especialmente reduzida na alternativa dos 90 km/h. O primeiro resultado poderá dever-se ao facto de esta alternativa ser normalmente atribuída a secções pertencentes a zonas de transição, que constituem as situações que envolvem, à partida, maior complexidade no processo de classificação, dada a sua natureza híbrida, com características mistas. Já o último caso poderá ser, mais uma vez, explicado pela menor ocorrência na base de dados de secções com atribuição de 90 km/h. Contudo, esta explicação é, por si só, insuficiente para justificar o desempenho particularmente fraco do algoritmo nesta alternativa, em que o número de secções atribuídas à categoria correta é apenas ligeiramente superior ao das classificadas com 70 km/h. De resto, esta é a única alternativa em que as árvores C4.5 apresentam um desempenho inferior ao das árvores CART, sendo este significativamente melhor nas restantes alternativas.

Em todo o caso, considera-se que, de um modo geral, o algoritmo com melhor desempenho global é o das árvores C4.5, tendo a árvore podada um maior poder explicativo que a não-podada. Esta situação é comum à das árvores do tipo CART, comprovando a eficácia do mecanismo. A árvore resultante da aplicação do algoritmo C4.5 após poda é apresentada na Figura 1.

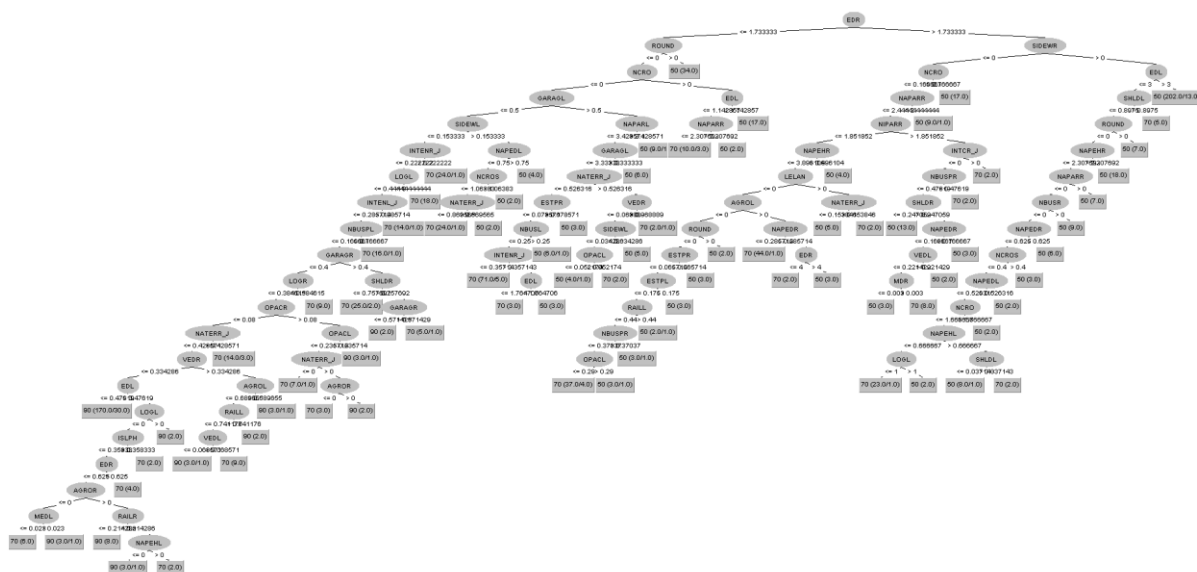


Figura 1. Configuração geral da árvore C4.5 podada (extraído do software WEKA 3.8)

A estrutura interna de uma árvore de decisão, ao nível da organização e disposição dos seus nós e partições associadas, não permite aferir diretamente a importância relativa dos atributos, pois baseia-se no subconjunto de instâncias considerado na construção de cada nó. A importância das variáveis independentes, avaliada através do seu contributo para a minimização geral da impureza, é apresentada no Quadro 5, sob a forma de uma lista ordenada contendo as variáveis mais significativas.

Quadro 5. Ordenação das variáveis pelo seu contributo para a redução da impureza (C4.5 podada)

Variável	Imp.	Variável	Imp.	Variável	Imp.	Variável	Imp.
SIDEWR	0,191	NAPEHR	0,130	GARAGL	0,103	NATERR	0,052
SIDEWL	0,183	NCRO	0,125	MEDL	0,099	NIPARL	0,048
EDL	0,175	MDL	0,109	AGROL	0,096	SHLDR	0,046
EDR	0,165	GARAGR	0,107	MEDR	0,095	NAPARL	0,043
NAPEHL	0,139	MDR	0,104	AGROR	0,095	NCROS	0,042

Os resultados apresentados por este indicador são, em geral, consistentes e intuitivos. De facto, as variáveis prevaletentes estão geralmente relacionadas com o tráfego pedonal (caso das variáveis SIDEW, representando a presença de passeios, ou NCRO, o número de passagens pedonais) ou com fatores característicos das zonas urbanizadas, tais como a existência de edificações (variáveis ED, MD e MED), de acessos pedonais privados ao edificado (NAPEH, correspondendo, na maioria das situações, a portas e portões) e de lugares de estacionamento privados (GARAG, que representa o caso clássico das garagens).

Tal como seria expeável, as variáveis representando a densidade de edificações (ED) encontram-se entre as mais significativas do modelo, o que corrobora o facto de os fatores relacionados com a densidade urbana – e, de alguma forma, com o nível de concentração e consolidação do padrão urbano – constituírem uma medida precisa e eficaz para a distinção entre áreas urbanas e rurais. Também por este motivo, seria de esperar que as variáveis relacionadas com a distância das edificações à via assumissem uma importância relativa mais significativa no modelo que aquela que acabaram por revelar. De facto, a consideração destes fatores pretendeu representar o grau de exposição das edificações à via (ao apenas ter em conta os elementos que formam parte da frente de edificado, em contacto direto com a faixa de rodagem), bem como a natureza da envolvente imediata. No fundo, procurou-se estabelecer se a via constitui um arruamento ou se, pelo contrário, a envolvente não contacta de forma direta com a estrada, independentemente da sua densidade. Este resultado poderá dever-se à grande heterogeneidade da malha urbana dos diversos aglomerados atravessados pelas vias incluídas na base de dados, compreendendo desde o típico caso de estrada-rua, até vias com um maior grau de segregação face à envolvente, mesmo em zonas urbanas mais consolidadas.

Por outro lado, seria de esperar que as variáveis representando o lado adjacente da via (terminação em R) assumissem uma maior relevância que as correspondentes ao lado oposto (terminação em L), dada a, em geral, maior sensibilidade do condutor a potenciais conflitos nesse lado. Porém, os resultados obtidos revelam uma situação mista, sem grande coerência ou padrão identificável. Este cenário poderá decorrer da metodologia empregue neste trabalho para a avaliação dos especialistas, que não decorreu através da observação das condições reais *in situ*, durante o ato da condução. De facto, as condições laboratoriais, mais controladas e uniformes, permitem abstrair o especialista da lógica de condutor, possibilitando uma avaliação mais estável e menos permeável ao efeito de condicionalismos momentâneos. Contudo, embora a metodologia seguida possa, igualmente, afastar o especialista das condições reais de circulação, não se considera que os resultados obtidos coloquem em causa a validade do método, dada a reduzida importância deste parâmetro e a muito limitada magnitude da diferença da importância das variáveis em causa, podendo inclusive apresentar o benefício de contribuir para uma maior homogeneidade do modelo final.

5 CONCLUSÕES

O presente trabalho, que está integrado numa linha de investigação mais vasta, pretendeu constituir uma metodologia de apoio à decisão para a definição do limite de velocidade em vias interurbanas de faixa de

rodagem única, em diversos tipos de ambiente rodoviário, e atravessando zonas urbanas, rurais e de natureza mista, em regiões com diferentes tipologias urbanas. Para tal, construiu-se um sistema especialista, tendo por base a avaliação de um grupo de especialistas, que analisaram um total de 304,2 km de vias, em ambos os sentidos, subdivididos em 1036 secções de extensão variável. Esta avaliação decorreu em laboratório, em condições controladas e homogéneas. A base de dados incluiu um conjunto de variáveis relacionadas com as características funcionais da via, do ambiente rodoviário e da envolvente, que foram recolhidas através de meios remotos. Para definição do algoritmo de decisão do sistema, foram testados dois tipos distintos de árvore de decisão: as árvores do tipo CART e o algoritmo C4.5, nas suas versões anteriores e posteriores a poda.

De um modo geral, os resultados obtidos pela aplicação do algoritmo C4.5 denotam um melhor poder explicativo destas árvores, apesar de o algoritmo CART não ser significativamente pior. Em ambos os casos, as árvores podadas resultaram num melhor desempenho que aquele das suas versões iniciais. As variáveis explicativas mais relevantes, apresentando um maior potencial de melhoria global da precisão, estão relacionadas com o tráfego pedonal, a densidade de edificações na envolvente e com outros fatores característicos das zonas urbanizadas, num resultado inteiramente intuitivo e expetável. Dados os pressupostos e a natureza destes algoritmos, não é possível extrair conclusões a partir da estrutura interna da árvore e da disposição dos seus nós e ramos.

Futuramente, poderão ser aplicados outros algoritmos envolvendo árvores de decisão, mas mais complexos e avançados. Um exemplo de um algoritmo deste tipo é o das Florestas Aleatórias (*Random Forests*), um preditor que recorre a um conjunto de árvores de decisão do tipo CART, introduzindo aleatoriedade na seleção dos subconjuntos de instâncias e de atributos utilizados durante o processo. Esta estratégia permite abordar a instabilidade e o sobreajustamento comuns nos algoritmos de árvore única. Por outro lado, numa fase futura, a metodologia deverá considerar segmentos de via mais extensos, agregando várias secções contíguas, de modo a analisar a consistência e a homogeneidade dos resultados, do ponto de vista do traçado e das transições entre secções, e contemplando então atributos relacionados com as características geométricas da via.

6 AGRADECIMENTOS

O primeiro autor foi financiado pela FCT – Fundação para a Ciência e a Tecnologia, I.P., ao abrigo do programa MIT-Portugal. Os autores agradecem ainda aos especialistas envolvidos neste projeto.

7 REFERÊNCIAS

1. G. Forbes, T. Gardner, H. McGee, R. Srinivasan, *Methods and Practices for Setting Speed Limits: An Informational Report*, Report No. FHWA-SA-12-004, Federal Highway Administration – Office of Safety, Washington, D.C., 2012.
2. J. Stuster, Z. Coffman, D. Warren, *Synthesis of Safety Research Related to Speed and Speed Management*, Publication FHWA-RD-98-154, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 1998.
3. L. Aarts, I. van Schagen, Driving Speed and the Risk of Road Crashes: A Review, *Accident Analysis and Prevention*, 38(2) 215-224, 2006.
4. E. Hauer, Speed and Safety, *Transportation Research Record: Journal of the Transportation Research Board*, 2103 10-17, 2009.
5. Department for Transport, *Setting Local Speed Limits*, DfT Circular 01/2013, Department for Transport, Londres, 2013.
6. FHWA, *Manual on Uniform Traffic Control Devices for Streets and Highways*, 2009 Edition, Federal Highway Administration, Washington, D.C., 2009.
7. TRB, *Managing Speed – Review of Current Practice for Setting and Enforcing Speed Limits*, TRB SR254, Transportation Research Board, National Academy Press, Washington, D.C., 1998.
8. Y. Park, F. Saccomanno, Evaluating Speed Consistency between Successive Elements of a Two-Lane Rural Highway, *Transportation Research Part A: Policy and Practice*, 40(5) 375-385, 2006.

9. R. Elvik, A Restatement of the Case for Speed Limits, *Transport Policy*, 17(3) 196-204, 2010.
10. R. Elvik, Can Injury Prevention Efforts Go Too Far? – Reflections on Some Possible Implications of Vision Zero for Road Accident Fatalities, *Accident Analysis and Prevention*, 31(3) 265-286, 1999.
11. C. Goldenbeld, I. van Schagen, The Credibility of Speed Limits on 80 km/h Rural Roads: The Effects of Road and Person(ality) Characteristics, *Accident Analysis and Prevention*, 39(6) 1121-1130, 2007.
12. H. Rosencrantz, K. Edvardsson, S. O. Hansson, Vision Zero – Is It Irrational?, *Transportation Research Part A: Policy and Practice*, 41(6) 559-567, 2007.
13. R. Elvik, Optimal Speed Limits: Limits of Optimality Models, *Transportation Research Record: Journal of the Transportation Research Board*, 1818 32-38, 2002.
14. Austroads, *Balance Between Harm Reduction and Mobility in Setting Speed Limits: A Feasibility Study*, Research Report AP-R272/05, Austroads, Sydney, 2005.
15. Minnesota Department of Transportation, *Methods for Setting Posted Speed Limits*, Transportation Research Synthesis, TRS 1204, Minnesota Department of Transportation, Saint Paul, 2012.
16. G. Correia, A. Bastos Silva, Setting Speed Limits in Rural Two-Lane Highways by Modeling the Relationship between Expert Judgment and Measurable Roadside Characteristics, *Journal of Transportation Engineering*, 137(3) 184-192, 2011.
17. A. Bastos Silva, A. Seco, N. Gregório, Setting Speed Limits in Interurban Single Carriageway Highways Using Experts Judgment, *Transport*, 31(2) 282-294, 2016.
18. N. Gregório, A. Bastos Silva, A. Seco, Speed Management in Rural Two-Way Roads – Speed Limit Definition Through Expert-Based System, *Transportation Research Procedia*, 13 166-175, 2016.
19. N. Gregório, A. Bastos Silva, A. Seco, Speed Management in Single Carriageway Roads: Speed Limit Setting Through Expert-Based System, in *Transport Infrastructure and Systems: Proceedings of the AIIT International Congress on Transport Infrastructure and Systems (Rome, 10-12 April 2017)*, (G. Dell'Aqua, F. Wegman, eds.), CRC Press, Leiden, 351–358, 2017.
20. S. Liao, Expert System Methodologies and Applications – A Decade Review from 1995 to 2004, *Expert Systems with Applications*, 28(1) 93-103, 2005.
21. P. W. Bonsall, H. R. Kirby, R. S. Kwan, *Applications of Expert Systems in Transport*, Working Paper 216, Institute of Transport Studies, University of Leeds, Leeds, 1986.
22. J. A. Wentworth, *Expert Systems in Transportation*, Technical Report WS-93-04, American Association for Artificial Intelligence, Palo Alto, 1993.
23. W. P. Wagner, Trends in Expert System Development: A Longitudinal Content Analysis of over Thirty Years of Expert System Case Studies, *Expert Systems with Applications*, 76 85-96, 2017.
24. R. Srinivasan, M. Parker, D. Harkey, D. Tharpe, R. Sumner, *Expert System for Recommending Speed Limits in Speed Zones*, Final Report, Transportation Research Board, Washington, D.C., 2006.
25. National Cooperative Highway Research Program, *An Expert System for Recommending Speed Limits in Speed Zones*, NCHRP Project 3-67, Research Results Digest, No. 318, National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C., 2007.
26. I. Moghimipour, M. Ebrahimipour, Comparing Decision Tree Method Over Three Data Mining Software, *International Journal of Statistics and Probability*, 3(3) 147-156, 2014.
27. G. G. Moisen, Classification and Regression Trees, in *Encyclopedia of Ecology* (S. E. Jørgensen, B. D. Fath, eds.), Elsevier Science, Oxford, 582-588, 2008.
28. L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, *Classification and Regression Trees*, Wadsworth Series, Chapman & Hall, Belmont, 1984.
29. J. R. Quinlan, *C4.5: Programs for Machine Learning*, Series in Machine Learning, Morgan Kaufmann, San Mateo, 1993.